

Kod szkolenia: **PENTAHO/DW**

Tytuł szkolenia: **Projektowanie hurtowni danych z wykorzystaniem pakietu Pentaho Business Intelligence**

Dni: 5

Opis:

Adresaci szkolenia:

Szkolenie jest adresowane do programistów, architektów oraz administratorów aplikacji, którzy chcą tworzyć lub utrzymywać hurtownie danych (DWH) z wykorzystaniem Pentaho Business Intelligence Suite.

Cel szkolenia:

Uczestnicy szkolenia zdobędą przekrojową wiedzę dotyczącą projektowania, implementowania, monitorowania, uruchamiania, strojenia procesów i elementów hurtowni danych, poznają ogólne założenia hurtowni danych tj. najpopularniejsze schematy relacyjnych hurtowni danych. Dzięki temu będą mogli wybrać właściwy zestaw narzędzi i technik dla swoich projektów. Szkolenie, poza ogólnym wprowadzeniem do pojęć teoretycznych, skupia się na stosie produktowym wybudowanym wokół Pentaho Business Intelligence Suite.

Mocne strony szkolenia:

Program obejmuje zarówno ogólne wprowadzenie w tematykę ETL, DWH, OLAP, jak i całościowe przedstawienie stosu produktowego Pentaho Business Intelligence. Szkolenie jest unikalne, gdyż tematyka poruszana w jego trakcie nie jest wyczerpująco ujęta w dostępnej literaturze, a wiedza na ten temat jest mocno rozproszona. Program jest ciągle uaktualniany ze względu na szybki rozwój rozwiązań, których dotyczy szkolenie.

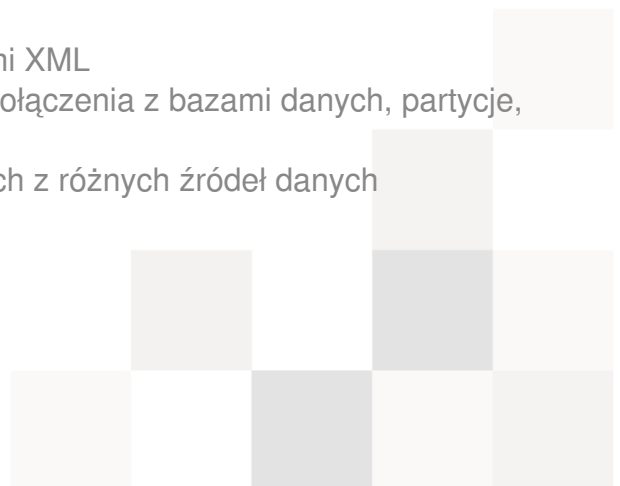
Wymagania:

Od uczestników wymagana jest podstawowa znajomość baz danych, podstawowa umiejętność programowania w języku Java.

Parametry szkolenia:

5*8 godzin (5*7 netto) wykładów i warsztatów, z wyraźną przewagą warsztatów. W trakcie warsztatów, oprócz prostych ćwiczeń, uczestnicy rozwiązują problemy przetwarzania danych, implementując własne procesy ETL, modelują struktury hurtowni danych, wykonują podstawowe czynności administracyjne.

1. Wprowadzenie do hurtowni danych i platformy Pentaho BI Suite
 - a. OLTP, OLAP, bazy danych, hurtownie danych, data marteny
 - b. ROLAP, MOLAP, HOLAP
 - c. Normalizacja, agregacja, fakty, wymiary
 - d. SQL, MDX, XML/A
 - e. ETL i integracja offline
 - f. BigData, BigTable, NoSQL, nierelacyjne hurtownie danych
 - g. Architektura i składowe Pentaho BI Suite
 - h. Business intelligence a hurtownie danych
2. Fizyczne struktury danych w hurtowni danych
 - a. Tabele faktów i tabele wymiarów
 - b. Indeksy, widoki, widoki zmaterializowane
 - c. ROLAP w szczegółach: schemat gwiazdy, płątka śniegu, konstelacji faktów
 - d. Testy wydajności ROLAP: The TPC Benchmark H
 - e. ROLAP vs MOLAP
 - f. Metody optymalizacji szczególnie przydatne w przypadku hurtowni danych
 - g. Wykorzystanie Pentaho Aggregate Designer do optymalizacji hurtowni danych
 - h. Znaczenie wymiaru czasu w hurtowniach danych
 - i. Dziedziczne hurtownie danych a dedykowane struktury danych np.: GeoSpatial DWH
3. Zasilanie hurtowni danych z wykorzystaniem Pentaho Data Integration (PDI)
 - a. ETL
 - i. Ekstrakcja danych
 - ii. Transformacja, oczyszczenie i uzupełnianie danych
 - iii. Ładowanie
 - iv. Wyznaczniki i ocena jakości danych
 - v. Stage'ing
 - vi. Hurtownie czasu rzeczywistego
 - vii. Kwestie wydajnościowe w kontekście procesów ETL
 - viii. Przegląd narzędzi ETL dostępnych na rynku
 - b. Pentaho Data Integration
 - i. Architektura
 - ii. Składowe
 - c. Praca ze Spoon
 - i. Instalacja, uruchomienie, Look&Feel (perspektywy, zakładki etc.)
 - ii. Zmienne
 - iii. Hop'y
 - iv. Praca z repozytorium i plikami XML
 - v. Współdzielenie obiektów tj. połączenia z bazami danych, partycje, klastry
 - vi. Czytanie i zapisywanie danych z różnych źródeł danych
 - vii. Transformacje
 - Walidacja danych
 - Uzupełnianie danych



- Transformacja danych
- Optymalizacja
- Wykorzystanie własnego kodu w transformatach
- Utilities
- Monitorowanie wydajności i analiza logów
- Wersjonowanie i praca grupowa
- viii. Zadania
 - Zadania (kjb) a transformacje (ktr)
 - Tworzenie złożonych zadań
 - Wykorzystanie własnego kodu w zadaniach
 - Przeływ pracy
 - Praca z plikami
 - Monitorowanie wydajności i analiza logów
 - Wersjonowanie i praca grupowa
- d. Praca z Kitchen i Pan
 - i. Uruchamianie zadań i transformat z plików i z repozytorium
 - ii. Scheduling
 - iii. Obsługa błędów
 - iv. Przekierowanie wejścia/wyjścia
- e. Praca z Cartle
- 4. Analityka z wykorzystaniem OLAP oraz Pentaho Analysis Services (Mondrian)
 - a. Definiowanie kostek OLAP na hurtowni danych
 - b. Pentaho Schema Workbench (PSW)
 - c. Model logiczny i model fizyczny
 - i. kostki
 - ii. miary
 - iii. wymiary, hierarchie, levele
 - iv. tabele
 - v. relacje
 - vi. agregaty
 - vii. wyrażenia
 - d. MDX w szczegółach
 - e. peracje Slice, Dice, Drill na kostkach
 - f. Optymalizacja na potrzeby OLAP
- 5. Raportowanie danych zgromadzonych w hurtowni danych z wykorzystaniem Pentaho Report Designer (PRD)
 - a. Definiowanie raportów za pomocą PRD
 - i. Źródła danych
 - ii. Modelowanie zapytań
 - iii. Wydobywanie danych
 - iv. Zawężanie i filtrowanie wyników
 - v. Wizualizacja danych na raporcie
 - b. Osadzanie raportów
 - c. Eksport do PDF, HTML, RTF etc.
 - d. Report Wizard
- 6. Raportowanie i analiza ad-hoc



- a. Pentaho BI Platform/Portal (BA Platform)
 - b. Pentaho Interactive Reporting (PIR)
 - c. Pentaho Metadata Editor (PME)
 - d. Pentaho Analityzer (PAZ)
7. Kokpity i dashboardy managerskie na danych zgromadzonych w hurtowni danych z wykorzystaniem Pentaho Dashboard Designer (PDD)
- a. Definiowanie dashboardów za pomocą PDD
 - i. Praca z wykresami - simple bar, line, area, pie, dial charts
 - ii. Praca z tabelami
 - iii. Praca z raportami
 - iv. Parametry
 - v. Szablony
 - b. Osadzanie dashboardów
8. Eksploracja danych zgromadzonych w hurtowni danych z wykorzystaniem Pentaho Data Mining (WEKA)
- a. Architektura
 - b. Praca z Explorer'em
 - c. Preprocessing i praca z danymi
 - i. Format danych ARFF
 - ii. Przygotowanie danych do analizy
 - iii. Odpowiedni dobór atrybutów np.: korelacja atrybutów a wyniki eksploracji danych etc.
 - iv. Filtrowanie i rodzaje filtrów w WEKA np.: filtrowanie, dyskretyzacja, normalizacja etc.
 - v. Wizualizacja
 - vi. Przetwarzanie dużych zbiorów danych, ograniczenia JVM 32bit
 - vii. Przetwarzanie strumieni oraz uczenie przyrostowe
 - d. Eksploracja danych
 - i. Klasyfikacja
 - ii. Grupowanie
 - iii. Odkrywanie reguł asocjacyjnych
 - iv. Ograniczanie i transformacja przestrzeni atrybutów
 - v. Pozostałe algorytmy i techniki eksploracji danych dostępne w WEKA
 - e. Rozbudowa możliwości WEKA
 - f. Wykorzystanie możliwości w połączeniu z innymi produktami Pentaho
9. Pentaho Mobile BI

